A study of music revolution based on influence network and similarity test

Wenchun Jing[1, *], Yaxuan Chen[1], Quanrun Qiu[2]

[1]School of Computer Science, NPU, ShaanXi, China, 710129

[2]Queen Marry University of London Engineering School, NPU, ShaanXi, China, 710129

*Corresponding author: vj162774@163.com

Abstract. This paper develops a method to quantify the evolution of music and understand the role of humans in the evolution of music. First, a directional music influence network was set to show the parameters of "music influence". Then, a sub-network of the direct influencer network was established to obtain influence relationships, and "musical influence" was described and stored in this sub-network. Finally, a music similarity test model is used to compare which is more similar between artists of the same genre and artists of different genres. By comparing the influence and similarity between genres, the difference and connection of genres was got. Analyze whether "influencers" can actually influence their artists and their music through the above-mentioned similarity data. Then analyze the influence of music characteristics. Identify features representing major evolutions in the development of music from the data and get influencers in the network that represent major evolutions; analyze the evolution of a musical genre over time and explain how the genre or artist has changed over time; and illustrate how the model Express the social, political or technological change at the time.

Keywords: Influence network, Visual analysis, Similarity test, Music genres

1. Introduction

The background of this question is to understand the development process of music based on the similarity and influence of music. Our task is to quantify and analyze this similarity and influence relationship, and then discover the influence relationship between influencers and followers, genres, and artists in the process of music development, discover the revolution in the music process[1].

First of all, for the relationship between influencers and followers, we have established an influence network, which is used to store influence relations and obtain influence parameters.

Next, conduct a similarity test. For this issue, the similarity test model we have established can compare a large number of artists. At the same time, after statistical analysis of the comparison results, a critical similarity coefficient can be obtained to distinguish whether they belong to the same genre. In specific operations, we found that the similarity of the same genre is significantly higher than that of different genres, and the critical similarity coefficient is 2.0.

To analyze this influence relationship, we also need to establish a visualization model. In this model, we visually analyze the number of artists, the number of works, and the characteristics of music, and get the progress of music development[2]. At the same time, with the help of the changes of music characteristics over time, we discovered the birth of revolution and traced the revolutionaries. With the help of the comparison of genres, we have successfully given the similarity test relationship of the genres.

2. Notations and Definition

Notations used in this paper is shown in Table 1.

Table 1. Notations used in this paper

Symbol	Description	Unit
Xi	Control group characters	/
Yi	Comparison group characters	/
Ai	Normalization constant	/
S	Similarity coefficient	/

Definitions mentioned in this paper is shown in Table 2.

Table 2. Definitions mentioned in this paper

Symbol	Description
Outdegree	The number of outward directed graph edges from a given graph vertex in a directed graph.
Indegree	The number of inward directed graph edges from a given graph vertex in a directed graph.
Transverse	It refers to one visit to each node in the tree (or graph) along a certain search route.

3. Model construction and solving

3.1. Music artist influence network model

The influence-data data shows more than 5000 music artists who have complicated influence and following relationships [3]. In this model, influence relationships are stored, and in-fluence parameters are determined. This section selects 1190 as the original influencer to obtain the network through his followers and follow-up followers (as shown in Table 3 above), conduct the model test, and determine the influence parameters.

Table 3. The influence relationship with 1190 as the original influencer

influencer_id	influencer_name	influencer_main_genre	follower_id	follower_name	follower_main_genre
1190	Syreeta	R&B;	5307	Alicia Keys	R&B;
1190	Syreeta	R&B;	41462	Solange	R&B;
1190	Syreeta	R&B;	163213	Tweet	R&B;
1190	Syreeta	R&B;	208774	Janet Jackson	R&B;
1190	Syreeta	R&B;	2660270	King	R&B;
5307	Alicia Keys	R&B;	165077	Stacie Orrico	Religious
5307	Alicia Keys	R&B;	370326	Sara Bareilles	Pop/Rock
5307	Alicia Keys	R&B;	997568	Karina	R&B;
5307	Alicia Keys	R&B;	2795896	Elle Varner	R&B;
5307	Alicia Keys	R&B;	3145706	Fifth Harmony	Pop/Rock
5307	Alicia Keys	R&B;	3407597	Astrid S	Pop/Rock
170770	Erykah Badu	R&B;	5307	Alicia Keys	R&B;
170770	Erykah Badu	R&B;	41462	Solange	R&B;
170770	Erykah Badu	R&B;	2660270	King	R&B;
208774	Janet Jackson	R&B;	5307	Alicia Keys	R&B;
208774	Janet Jackson	R&B;	262255	Mariah Carey	R&B;
208774	Janet Jackson	R&B;	422891	Nivea	R&B;

To solve the follow-up problems, firstly an influence network of musicians was established. Building a relationship between influencers and followers relies on the depth-first search algorithm (DFS) to obtain the transfer relationship of influence. Since an influencer has several followers, and followers can be influencers of other followers, so transfer relationships ultimately form a transfer network[4].

In this model, each musician is stored as a node, and the directional arrow points to the following artist. The out-degree reflects influence parameters, and the presence or absence of the in-degree determines to represent it as the original influencer. After the DFS algorithm determines the node and the out-degree in-degree, use *Csacademy* to draw the network, from which a subset of music influence can be obtained, and the "influence parameters" of each influencer can be obtained.

The influence network with 1190 as the original influencer is shown in Figure 1.

The influence parameter of this network calculated by model is shown in Table 4.

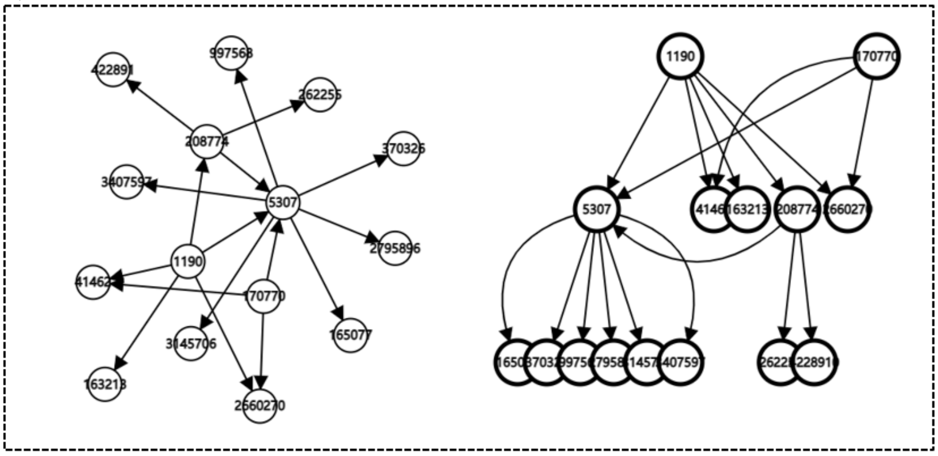


Figure 1. The influence network with 1190 as the original influencer

Table 4. The influence parameter of this network calculated by model

Influence paremeter							
Influencer	*1190*	*170770*	*5307*	*41462*	*163213*	*208774*	*2660270*
Influence	*5*	*3*	*6*	*0*	*0*	*2*	*0*

Just like the result show, between five followers of *Syreeta*, *Alicia Keys* owns clear advantages compared to other followers *Solange Tweet Janet Jackson and King. She owns more works, more awards, and more public honors*.

After the influence relationship network is established, the network is ready to be applied to analyze the similarity and influence relationship in the follow-up model.

3.2. Construction and solution of similarity test model

Suppose there are two artists X and Y, x*i* and y*i* represent their 11 groups of characteristics, and the absolute value of each characteristic difference m*i* is stored in another matrix M. Each characteristic corresponds to a normalized coefficient a*i*, which is approximately the reciprocal of the maximum range, *ai* are stored in matrix A[5]. The similarity coefficient of artists is approximately the product of two matrices. Similarity test model data is shown in Table 5.

Table 5. Similarity test model data

artist	danceability	energy	valence	tempo	loudness	mode	key	acousticness	instrumentalness	liveness	speechiness
X	*x1*	*x2*	*x3*	*x4*	*x5*	*x6*	*x7*	*x8*	*x9*	*x10*	*x11*
Y	*y1*	*y2*	*y3*	*y4*	*y5*	*y6*	*y7*	*y8*	*y9*	*y10*	*y11*
M	*m1*	*m2*	*m3*	*m4*	*m5*	*m6*	*m7*	*m8*	*m9*	*m10*	*m11*
A	*a1*	*a2*	*a3*	*a4*	*a5*	*a6*	*a7*	*a8*	*a9*	*a10*	*a11*

$$Mi = | xi - yi | \tag{1}$$

$$Ai = 1 / | \mathrm{Max} - \mathrm{Min} | \tag{2}$$

$$S = A \cdot M \tag{3}$$

The similarity coefficient calculated by this formula reflects the similarity degree of many artists. The smaller the value is, the higher the similarity degree is. If the differences are similar, and if there is a similar relationship between artists and genres, A statistical analysis based on a large number of experiments has to be done, by which the boundary coefficien S that can judge whether there is similarity can be obtained.

3.2.1 Judging genre by the similarity through music artists

To compare the similarity, the critical coefficient is required first. By using the model, 200 music artists from the same genre and different genres are compared with the control group (only 10 groups of data are listed here)[6]. We find that the similarity coefficient of music artists from the same genre is within the range of (0.2.0), while those beyond this range are artists from different genres. It can be seen that the similarity coefficient of artists of the same genre is significantly higher than that of artists of different genres. At the same time, through this analysis based on statistics, finally determined that the critical coefficient of testing whether music artists belong to the same genre is 2.0. Artists from the same genre and different genres are compared with the control group (examples) is shown in Table 6.

Table 6. Artists from the same genre and different genres are compared with the control group (examples)

artist_id		danceability	energy	valence	tempo	loudness	mode	key	acousticness	instrumentalness	liveness	speechiness	similarity coefficient
183867	*control group*	0.461357143	0.0798093	0.18695	101.778	-26.06007143	1	5	0.985071429	0.776214286	0.1244	0.066142857	
	The same genre												
183867		0.461357143	0.0798093	0.18695	101.778	-26.06007143	1	5	0.985071429	0.776214286	0.1244	0.066142857	0
239859		0.323771904	0.2117412	0.23865176	105.145	-19.34347505	1	5	0.87542329	0.506012168	0.227327357	0.085401294	1.04191978
37658		0.569	0.3485	0.4645	117.355	-16.049	1	7	0.8815	0.5425	0.09615	0.04865	1.590932992
849672		0.278972222	0.1099417	0.09781111	110.108	-23.20713889	1	5	0.914416667	0.831147428	0.128794444	0.040933333	0.607394733
579458		0.256142857	0.2410024	0.17509286	101.529	-15.35130952	1	2	0.939	0.040716155	0.276147619	0.047002381	1.922534738
641626		0.29625	0.105005	0.1467	125.021	-22.767	1	4	0.984	0.049023	0.09255	0.06295	1.350168267
929776		0.23675	0.0194068	0.1042	78.051	-33.47625	1	5	0.82675	0.0925706	0.09545	0.070025	1.619219551
731975		0.31860274	0.2058658	0.19913425	104.745	-16.3669589	1	2	0.976315068	0.054977048	0.294519178	0.073182192	1.763721795
151666		0.397866667	0.1107733	0.32933333	105.025	-24.42966667	1	4	0.9552	0.8782	0.105906667	0.048586667	0.574924441
803752		0.359996871	0.2222024	0.64384188	110.241	-22.1191444	1	7	0.980315283	0.866343671	0.139213117	0.047417329	1.197291254
	Different genre												
18513		0.7348	0.7	0.8468	128.022	-12.2554	1	11	0.05377	0.538838	0.19374	0.0645	3.913538311
830028		0.735	0.7996667	0.941	126.716	-6.967	1	3	0.629	0.119150513	0.228766667	0.059433333	3.764648432
409985		0.735	0.521	0.946	113.545	-12.161	1	1	0.655	0.024	0.207	0.0317	3.546073674
867980		0.742409836	0.7397377	0.72167213	133.342	-4.999819672	0	10	0.213640984	0.0000498	0.16085082	0.11104918	5.34891647
69402		0.742844884	0.5476122	0.74933498	120.299	-9.665752475	1	9	0.204243982	0.028482358	0.289808581	0.153062046	4.066676447
787940		0.742875	0.4614375	0.5865625	119.175	-10.3906875	0	5	0.2148625	0.000431913	0.10545625	0.063225	4.195773791
624215		0.744	0.6156316	0.73910526	109.826	-12.76526316	1	11	0.321936842	0.292993105	0.075321053	0.056194737	3.575392301
2385504		0.75212381	0.725419	0.65625714	124.864	-4.849952381	0	11	0.137885714	0.009048336	0.172227619	0.14337619	5.433137872
116425		0.7522	0.7624	0.6298	116.579	-5.3414	0	8	0.093394	0.013948	0.15928	0.07602	5.06399132
2528416		0.752333333	0.6632222	0.47811111	115.834	-6.960222222	1	0	0.069686667	0.070983694	0.178138889	0.109877778	3.96455147

3.2.2 Test of Critical Coefficient of Similarity

In order to test the reliability of the critical coefficient, all *72* followers of influencer *Bob Dylan* were selected, and Bob Dylan was used as the control group for similarity test. The proportion of followers whose similarity coefficient is in the range of [0,1] [1,2] [2,3] is 33%, *57%* and *10%* respectively. *90%* of the followers are in the critical coefficient, which proves the reliability of this test method. All Bob Dylan followers were compared with Bob Dylan to test Critical Coefficient is shown in Figure 2.

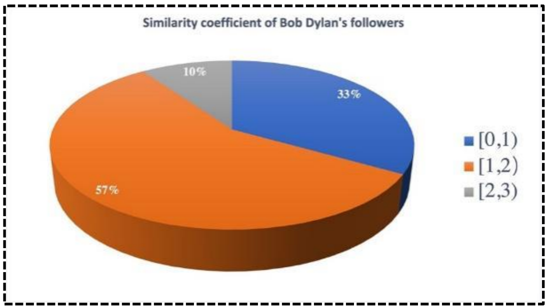


Figure 2. All Bob Dylan followers were compared with Bob Dylan to test Critical Coefficient

3.2.3 Prove influence relationship through similarity analysis

After the above influence network was established, the relationship of influencers and followers was stored. In order to analyze whether their relationship of influence will bring similar musical characteristics,1190 and his first- generation and second-generation followers were chosen to conduct a similarity test. The results are shown in the figure below. It shows that in the comparison between followers and influencers, the similarity coefficient is mostly less than the critical value of 2.0. Supporting the similarity data obtained from our model can indicate that the influencer is actually influencing the corresponding artist.

3.2.4 Similarity test among different genres

In order to compare the similarities between different genres, we compare the characteristics of different genres one by one. This process is similar to the comparison between the above artists. However, due to the limited number of genres and the similarity coefficient are limited to (0.3), it is difficult to get the basis of similarity judgment. Only through extreme value judgment can we get the relationship of high similarity and low similarity.

This problem can be solved synthesizing the next model.

3.3. Visual analysis model

The purpose of this model is to realize the visual analysis of data influence relations and describe the development of music by introducing time[7]. Similarly, influence, various music characteristics, and genres can all be created to visualize line graphs, which can help analyze actual problems.

Firstly, the data structure used for temporary storage in memory is established. Next, it is decided whether to traverse the *influence_data* file as needed, extract the genre list, and establish data filtering and classification functions according to the functional requirements.

Then, open the main file, remove the data other than the available data in the file, traverse the main file one or more times, use the filter and classification function to load the data in the file into the memory, and process the data in the memory.

By comparing the data sets in *influence_data* file and *full_music_data* file, the intersection of them is obtained. If the intersection is large enough, the image is generated. Otherwise, the data filtering and classification functions are modified. Generate the corresponding image on demand and check the image visualization. If the visualization is good, output the image. Or modify the data processing method in memory[8].

Analyze the image data, compared with the original data set, through the sampling method to verify the rationality of the image. If it is reasonable, output the analysis result. Or modify the processing method of data in memory. A realization process of visual analysis curve is shown in Figure 3.

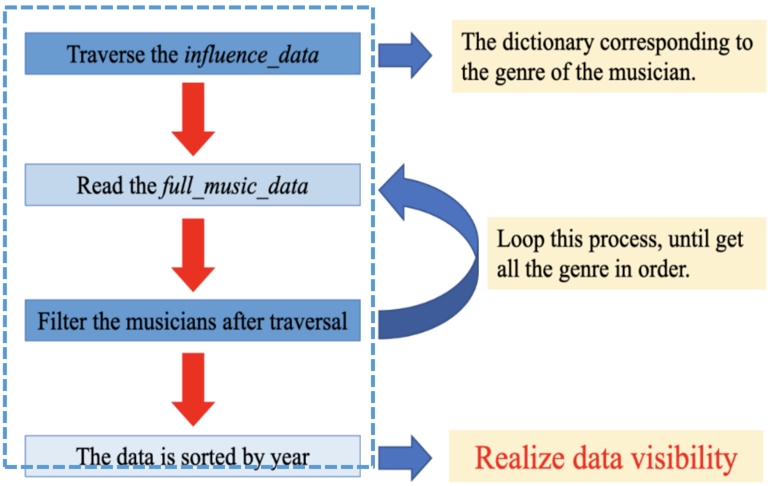


Figure 3. A realization process of visual analysis curve

3.4. The solution of visual analysis model

In this part, we get the influence of genres, the number of genre songs, music characteristics, the number of genre artists, and the popularity of genre over time[9]. In particular, the last picture, because of the need to express the number of songs and the popularity of each song at the same time, chose to make a scatter plot, the Figure 4-7 show us the development trend of the genre visually.

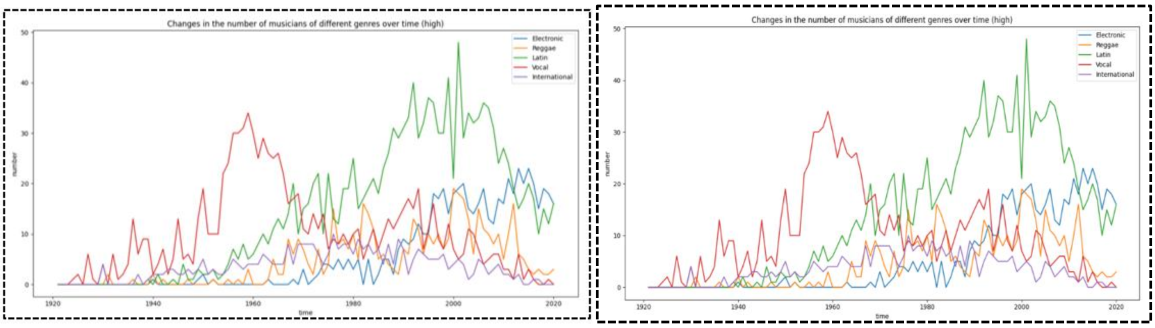


Figure 4. The number of artist over time

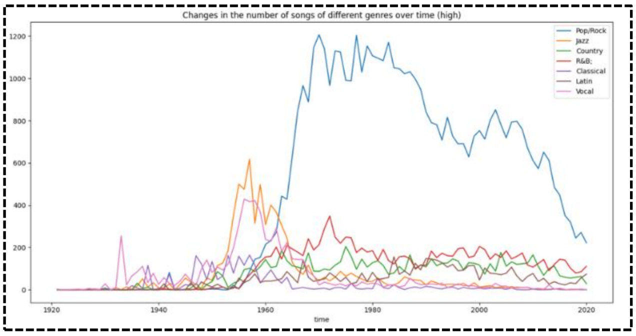


Figure 5. The number of song over time

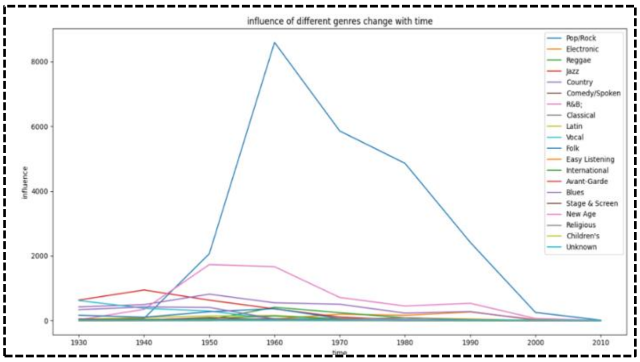


Figure 6. The influence of different genres over time

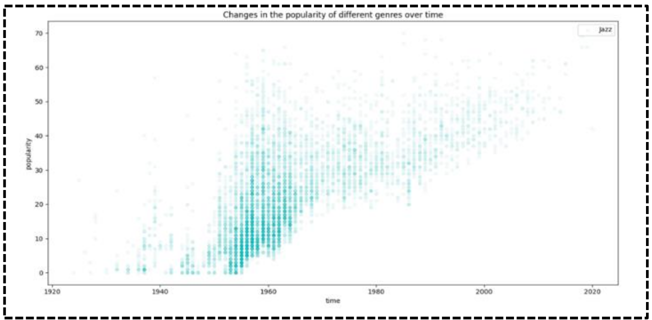


Figure 7. Scatter plot with “popularity” and“ number of songs”

From the figures above, we can see that before the Second World War, the vocal genre was very popular, but then from around 1950, the number of musicians decreased, and the roll and Latin genre increase. This also reflects that there is a connection between different genres. The number of musicians reflects the rise and decline of the genre. At the same time, since the 1950s, the rock music genre has been far ahead of other genres in terms of music number and artist number and has risen rapidly, which also corresponds to the curve of genre influence and time. In recent years, the rock genre has gradually returned to normal development, which can be found through the influence curve, the number of songs, and the number of artists curve.

For scatter plot, it can reflect the number of songs through the change of points, and its density reflects the popularity. The Jazz genres reached its peak before 1960, which can be con-firmed on several figures. At the same time, in recent years, this genre has been declining, this also shows a trend of reducing the number of points and low density on the scatter diagram.

Through these figures, we find that the model can describe the development of genres. At the same time, different analysis curves can complement each other and prove the correct-ness of the analysis.

This part shows Figure 8-10 they are *"Loudness" "Popularity"* and *"Speechiness"* change over time. Just like the figures show, we establish the relationship between characters with the time to understand the major evaluation[10].

- In the Loudness over time figures, in the 1980s there exists a large increase, we can find the musicians and the corresponding songs, they are *"Molly Hatchet" "The Clash" "Roxy Music"* and *"Electric Light Orchestra"* and the songs are *"Double Talker" "Version City – Remastered" " Oh Yeah!" "I'm Alive"*, these artists can be seen as revolutionaries.

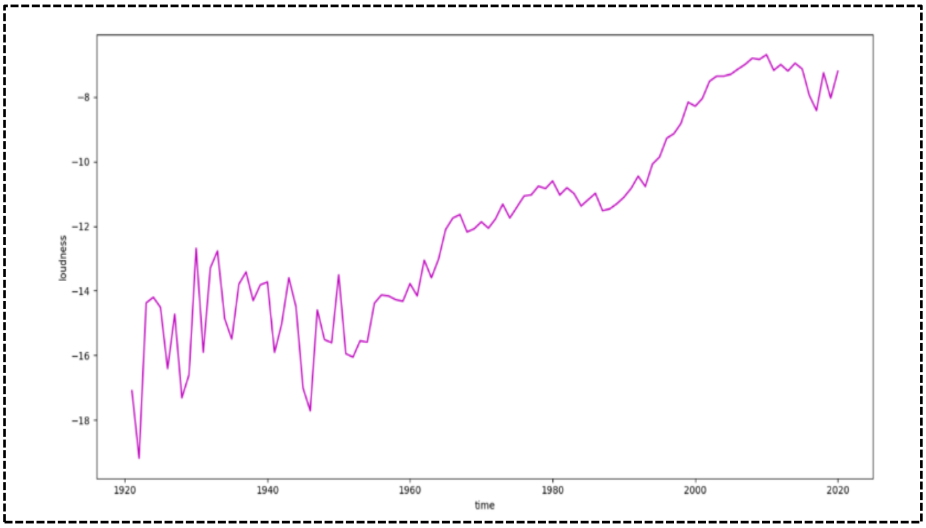


Figure 8. Loudness influence over time

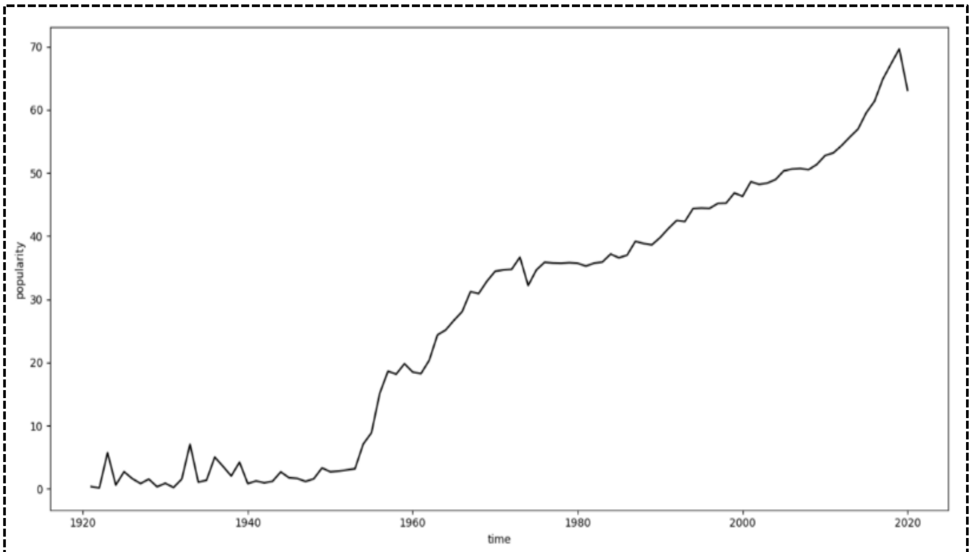


Figure 9. Popularity influence over time

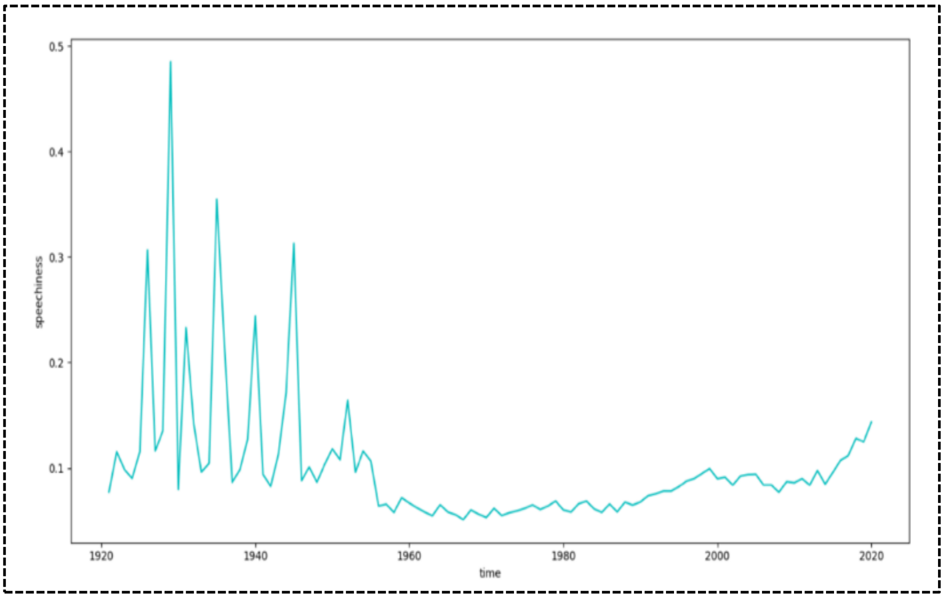


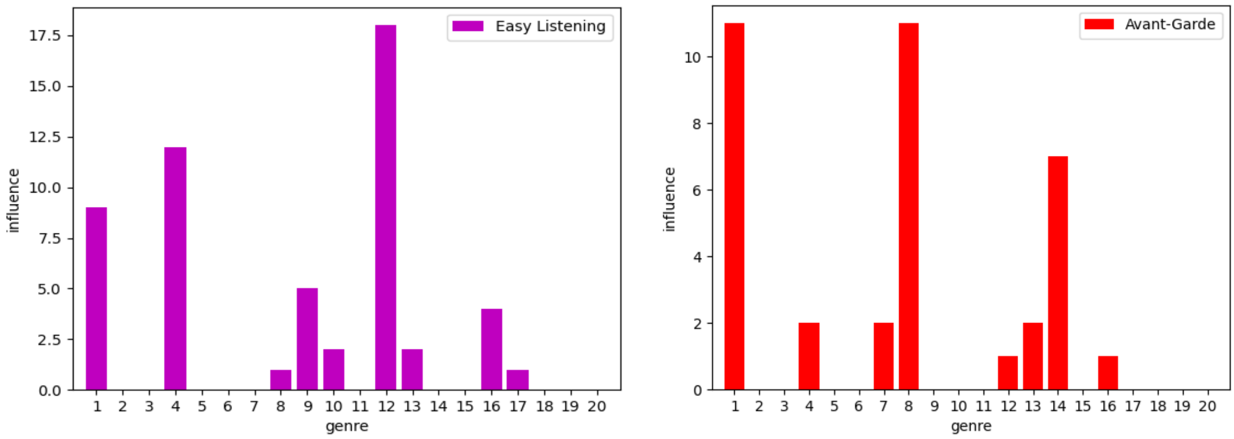
Figure 10. Speechiness influence over time

- Similarly, near 2020, popularity influence reaches the peak, and the revolutionaries are *"5 Seconds of Summer" "Keith Urban" "Kygo"* and *"Sam Hunt"*.
- About 1960 Spechiness influence strat to decline we find that *"Jerry Lee Lewis"*.
- *"The Everly Brothers" Conway Twitty" "Betty Carter"* are revolutionaries, their parameters of speechiness are extremely low which results in the decline.

This model can visually analyze the changes of certain characteristics, and at the same time, we can find revolutionaries through the network and can verify the correctness of this inspection method.

- The similarity test of the two genres

Since the previous similarity test method is not suitable for the comparison between genres and genres, the final solution of this problem is solved by establishing a visual model between genres and genres. Each genre is affected by other genres. But this kind of influence can be small or large. The biggest one must be oneself, and besides oneself, the relationship that has a larger influence is what we are looking for. Due to space reasons, this part will not be described in detail, and pictures were listed for explanation. The similarity test of the two genres is shown in Figure 11.



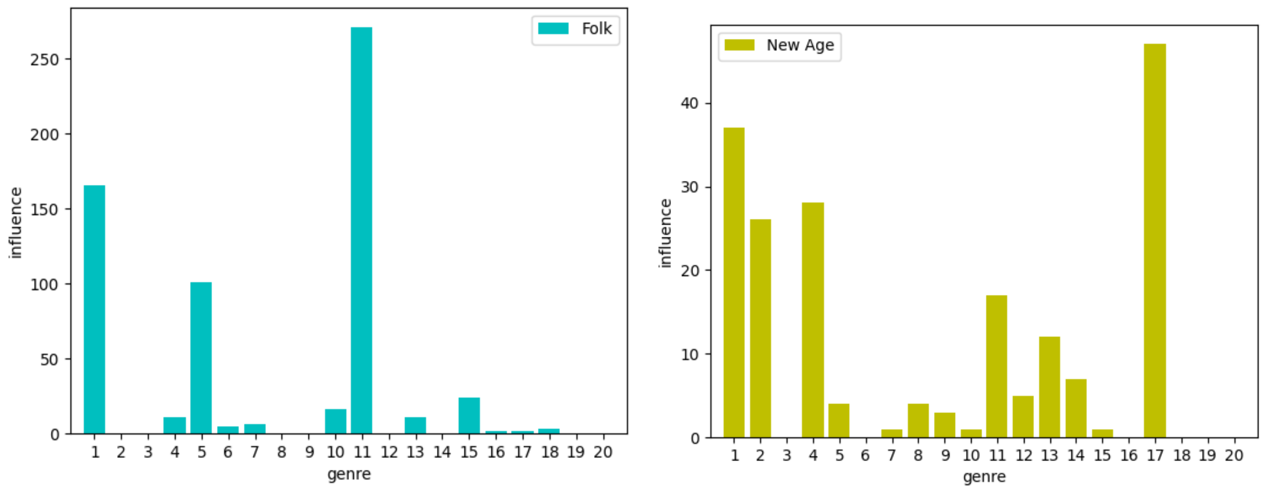


Figure 11. The similarity test of the two genres

4. Conclusion

To find the influencing relationships, Influence network model, Similarity test model and Visual analysis model were built. By these three models, we stored the relationship between influencers and followers, compared the similarities of artists, found that the same genre is more similar. In addition, we analyzed the characteristics of music visually according to the visual model, got the evolution process of the genre, and found the main changes and revolutionaries. Finally, the model can be further extend and modify. Through these three models, we have completed the exploration of music evolution and understood the role of human factors.

References

[1] Bergstra J, Casagrande N, Erhan D, et al. Aggregate features and AdaBoost for music classification[J]. Machine learning, 2006, 65(2): 473-484.

[2] Georges P. Western classical music development: a statistical analysis of composers similarity, differentiation and evolution[J]. Scientometrics, 2017, 112(1): 21-53.

[3] Wang X, Kang Y, Zhang Y, et al. Analysis and Modeling of the Music Influence and Evolution[C]//2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR). IEEE, 2021: 152-162.

[4] Simonetta F, Carnovalini F, Orio N, et al. Symbolic music similarity through a graph-based representation[M]//Proceedings of the M Audio ostly 2018 on Sound in Immersion and Emotion. 2018: 1-7.

[5] Berenzweig A, Logan B, Ellis D P W, et al. A large-scale evaluation of acoustic and subjective music-similarity measures[J]. Computer Music Journal, 2004: 63-76.

[6] Dan Shi. Research on music style similarity detection algorithm[D]. Dalian University of Technology,2013.

[7] Patrick George, Nguyen Nguyen, Zhang Jia-Ming. Visualizing musical similarity: clustering and mapping of 500 classical music composers[J]. Digital Humanities Research,2022,2(01):68-85.

[8] Shen Guoming. Design and implementation of a music recommendation system based on original tags [D]. Beijing University of Posts and Telecommunications.

[9] Aucouturier J J, Pachet F. Representing musical genre: A state of the art[J]. Journal of new music research, 2003, 32(1): 83-93.

[10] Mauch M, MacCallum R M, Levy M, et al. The evolution of popular music: USA 1960–2010[J]. Royal Society open science, 2015, 2(5): 150081.